The limits and desirability of automated data analysis

Jason Leman Surveys Officer Higher Education Academy



Session summary

- >Can Institutional Research be automated?
- >Can all kinds of data analysis be automated?
- > A practical example
- >Implications for Institutional Research

The electric researcher is born

Probability that statisticians will be automated: 22%

"Develop or apply mathematical or statistical theory and methods to collect, organize, interpret, and summarize numerical data to provide usable information."

Probability that survey researchers will be automated: 23%

"Plan, develop, or conduct surveys. May analyze and interpret the meaning of survey data, determine survey objectives, or suggest or test question wording" Probability that mathematical technicians will be automated: 99%

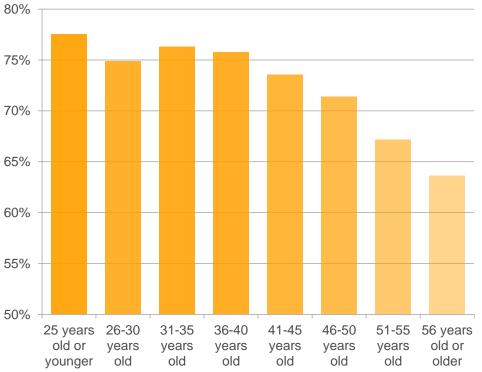
"Apply standardized mathematical formulas, principles, and amethodology to technological problems in engineering and physical sciences in relation to specific industrial and research objectives, processes, equipment, and products"

(Frey and Osbourne 2013)

> Descriptive analysis...

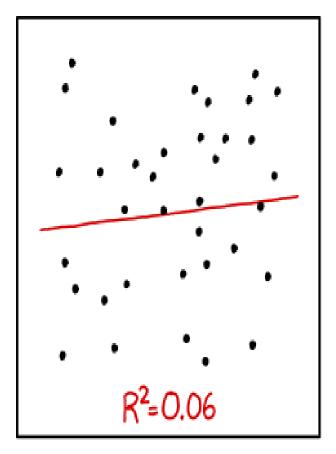
- Descriptive analysis is often the first step of analysis, setting out broad relationships in the data
- Automation of analysis often starts with descriptive analysis
- Without context, the meaning of these relationships may be mis-understood
- Without statistical testing, spurious differences may be given credance

Encouraged to think about devoloping career skills



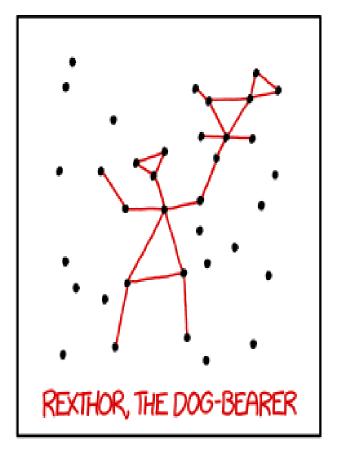
Regression and modelling

- Regression analysis may be used to evaluate the impact of factors upon particular measures
- Survey research has an issue in that factors are often closely correlated, which makes regression analysis less reliable in estimating true relationships
- Machine modelling of data is sensitive to how variables are described and used - garbage in, garbage out

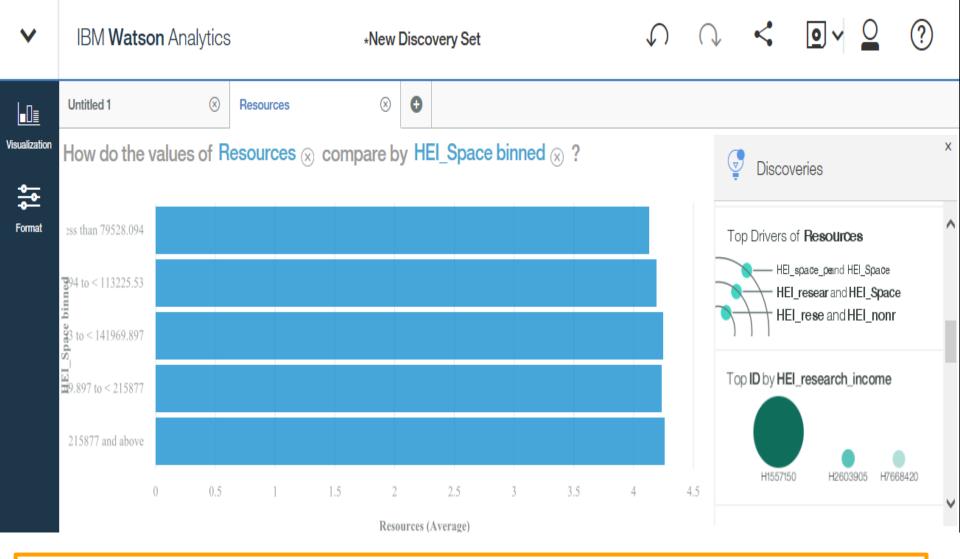


Regression and modelling

- Regression analysis may be used to evaluate the impact of factors upon particular measures
- Survey research has an issue in that factors are often closely correlated, which makes regression analysis less reliable in estimating true relationships
- Machine modelling of data is sensitive to how variables are described and used - garbage in, garbage out



(XKCD)



Watson Analytics enables sophisticated analysis of data 'at the push of a button', but does not inform the user as to the research process that underlies that analysis. The user is presented with an 'answer', rather than with a nuanced picture of uncertainty.



The tool is useful for the easy exploration and visualisation of data, but the user needs awareness of what such relationships mean to draw robust conclusions

eyondCore also offer analysis 'at the touch of a button', but in the smallprint make clear that thout a good understanding of the data, results are likely to be meaningless. Whilst these ickages are promoted as needing little expertise, there is a risk of 'garbage in – garbage ou

Rows represent Transactions

and the unit is Units

(Purchases, Patient Visits, Transactions, ...)

(Dollars, Days, Returns, ...

12

What do the rows represent?

BeyondCore only knows what is in your data, so it doesn't know what each row represents. Each row might be a customer, a patient visit, a purchase, a shipment, etc. Here you specify what BeyondCore should call each row in the story.

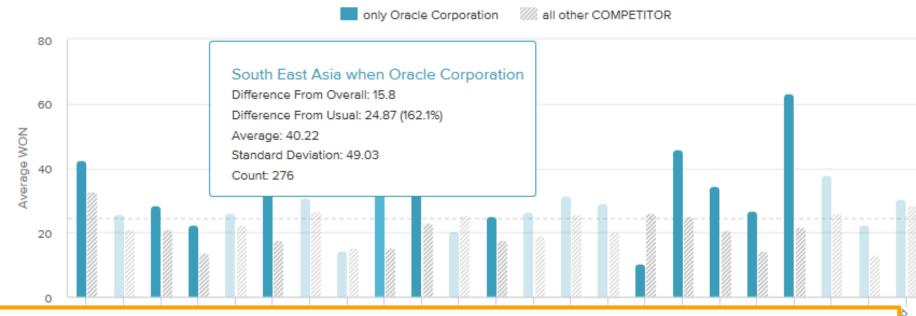
A When you upload your own data to BeyondCore, be very careful about the structure of your data. Each row should be at the right granularity for your analysis. For example, if I want to analyze Customer Lifetime Value, each row should be a customer, and the outcome would be the total amount of purchases by the customer. But if I wanted to understand the pattern of purchases in general, each row would be a purchase and the variable I want to understand would be the purchased amount.



D

WINLOSS

WON BY REGION WHEN COMPETITOR IS ORACLE CORPORATION



As with Watson, BeyondCore offers good options around the visualisation of data. There is also some contextual information on statistical tests and pointers towards analytical strategies.

When COMPETITOR is Oracle Corporation, REGION: Midwest and Channel do better

But you should also be aware of the case when COMPETITOR is Oracle Corporation, because the WON demonstrated a different pattern in this case. If you compare the two graphs, you

> Factor analysis and SEM...

Factor analysis is used to simplify analysis and dissemination through exploring higher level factors. Multiple items may be combined into scales and scalelets that are typically more reliable as measures than individual items.

Factors are often built into questionnaire design, in which case Structural Equation Modelling is more appropriate. However, SEM is more complex, and thus is typically underused. "Exploratory factor analysis (EFA) is rightly described as both an art and a science, where researchers follow a series of analytic steps involving judgments more reminiscent of qualitative inquiry, an interesting irony given the mathematical sophistication underlying EFA models."

Osborne et al (2008)

> Factor analysis...

	Component		Component			Com		onent	
	1	2	1	2		1	2	3	4
Q16.a Writing	0.179	0.631	0.72	0.013		-0.059	-0.013	-0.011	-0.919
Q16.b Speaking	0.336	0.507	0.621	-0.129		0.235	-0.142	0.004	-0.449
Q16.c Thinking	0.12	0.697	0.778	0.065		0.26	-0.016	0.207	-0.365
Q16.d Analysing	-0.226	0.789	0.511	0.065		-0.013	-0.022	0.682	-0.021
Q16.e Employability	0.053	0.645	0.531	-0.054		0.258	-0.082	0.344	-0.049
Q16.f Independent	0.189	0.561	0.6	-0.045		0.507	0.009	0.119	-0.116
Q16.g Innovative	0.405	0.388	0.499	-0.201		0.767	-0.012	-0.047	-0.016
Q16.h Working	0.459	0.314	0.402	-0.283		0.514	-0.183	0.051	0.01
Q16.i Developing	0.826	0.032	0.087	-0.735		0.207	-0.657	-0.065	-0.019
Q16.j Understanding	0.912	-0.119	-0.1	-0.876		0.003	-0.825	-0.11	-0.03
Q16.k Exploring	0.74	0.084	0.121	-0.654		-0.043	-0.724	0.167	0.003
Q16.I Active	0.807	0.051	0.071	-0.752		-0.026	-0.783	0.069	-0.036
	PCA		PAF		PAF - extracted four fac				S

Types of analysis like factor analysis require not just statistical techniques, but a research approach that incorporates judgements around what is meaningful and desirable. Automated analysis may make such processes easier, but need to enable informed research choices rather than presenting results as concrete.

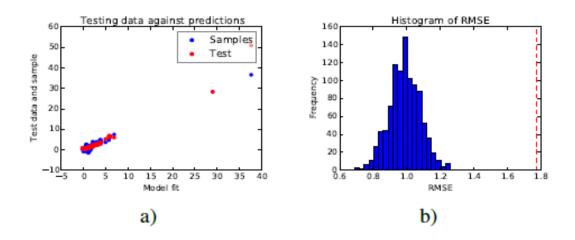
4 Model criticism

1

4

In this section I have attempted to falsify the model that I have presented above to understand what aspects of the data it is not capturing well. This has been achieved by comparing the model with data I held out from the model fitting stage. In particular, I have searched for correlations and dependencies within the data that are unexpectedly large or small. I have also compared the distribution of the residuals with that assumed by the model (a normal distribution). There are other tests I could perform but I will hopefully notice any particularly obvious failings of the model. Below are a list of the discrepancies that I have found with the most surprising first. Note however that some discrepancies may be due to chance; on average 10% of the listed discrepancies will be due to chance.

Systems such as "The Automatic Statistician", now funded by Google, raise the possibility of advanced analysis, standardised research strategies, and text outputs. Here, the programme has created a paper that describes research choices and statistical uncertainty in plain English.



Lloyd (2014) The Automatic Statistician

- There is increasing competition for "easy to use" analytical software, with statistics program moving to cater for easier analysis
- Programs like IBM Watson analytics offer analysis "at the push of a button", guiding automated data analysis through machine learning and statistical analysis, including factor analysis, neural networks, and regression
- Proposals for student surveys that automate factor analysis to produce "on the fly" scales

- There is increasing potential for smart data analytics, offering sophisticated data analysis alongside embedding multiple kinds of evidence alongside quantitative measures
- > Analysis "at the push of a button" risks poor decisions being made
- Good smart analytics programmes guide users through research choices and inform them of the fuzziness of findings

Making my own electric researcher

	tu Martin												
	7663 Compo Inglio Man (* 227 48 7663 Compo Inglio Card a 216 36 7663 Compo Inglio Card a 216 36 7663 Compo Inglio Malena 184 19 7663 Compo Inglio Malena 23 1 7663 Compo Inglio Malena 3	A 267 266 267 26 17 86 98 81 81 16 118 128 122 11 17 88 28 43 41 17 88 28 43 41 18 8 18 16 16	C 002 002 022 5 366 360 360 1 09 110 94 1 02 110 127 1 42 33 40 1 7 6 3	012 242 022 262 266 260 267 267 43 26 43 80 43 40 28 44 2 13 8 12	bit bit <td>MIX BXX BXX B4X B4X Sun Sun</td> <td></td> <td></td> <td>022 002 022 022 024 024 024 024 024 024</td> <td>14 00 07 040 140 : 13 117 131</td> <td>7% 62% 13% 11% 48% 83% 68 368 368 368 368 368 23 183 334 334 334 148 136 </td> <td>232 132 342 340 340 70 1 354 354 354 354 354 354 3 174 323 141 388 344</td> <td>112 382 172 87 172 112 112 112 112 112 112 112 112 11</td>	MIX BXX BXX B4X B4X Sun Sun			022 002 022 022 024 024 024 024 024 024	14 00 07 040 140 : 13 117 131	7% 62% 13% 11% 48% 83% 68 368 368 368 368 368 23 183 334 334 334 148 136 	232 132 342 340 340 70 1 354 354 354 354 354 354 3 174 323 141 388 344	112 382 172 87 172 112 112 112 112 112 112 112 112 11
	JUERS Compaction & Month 402 41 JUERS Compaction & Control 3440 3443 JUERS Compaction & Database 10003 1044 JUERS Compaction & Montro 30100 1044 JUERS Compactions & Montro 30100 1044 344	2 872 882 882 882 783 8 34966 34988 34889 3486 17 18338 16866 13688 9486 14 14688 13388 16386 1388 14 3438 3878 4833 6835	4 34487 34836 34877 34 7 11441 16136 13847 14 8 14686 13411 13674 13 4 8134 3466 8168 4	4463 34437 34884 34871 34 4366 8443 13481 11388 1 3433 16483 14178 14374 14 4446 8784 8374 4244 1	1307 34440 34343 34400 36464 3 1136 0401 10306 37040 13646 1633 13474 14333	1348 34999 33999 33881 3478 2001 8398 1368 10138 1008 1378 9793 2099 3233 1638 3088 3961 3289 3629 446	1 34361 34440 34000 34044 1 12634 10440 13031 4634 1 14634 10023 14006 13313 2003 4400 4737 7600	24626 24816 22227 22618 2 18748 16681 14884 12482 12428 14864 12668 12684 2828 2424 2244 4626	4876 34811 34634 34688 3444 14188 11833 14844 11331 1148 14341 13717 13888 14348 1331 4688 6844 4383 6858 863	X 70X 04X 40X 00X 1 4 34077 34401 30067 30067 30 8 13076 13010 33030 10167 30 11 13076 13010 33030 10167 30 11 13013 16343	167 20067 20067 20067 20067 20067 121 10664 20447 22007 17646 10602	30067 30067 30067 30067 30067 30067 30 10647 34373 30046 30134 3004 33	20 40 40 40 40 40 40 40 40 40 40 40 40 40
	ALCES Fragma from Cr. Briefen (* 276) 21 ALCES Fragma from Cr. Minner (* 443) 1.43 ALCES Fragma from Cr. Minner (* 443) 1.43 ALCES Fragma from Cr. Minner (* 414) 1.43 ALCES Fragma from Cr. Minner (* 416) 1.43 ALCES Fragma from Cr. Minner (* 416) 1.43	14 427 664 623 628 14 6.046 6.407 6.407 6.2 17 6.666 6.426 6.444 6.2 18 6.667 6.644 6.424 6.42 18 6.621 6.607 6.444 6.444 18 6.621 6.622 6.216 6.43	7 1214 282 776 4 8.842 8.482 8.488 8 7 8.777 8.876 8.782 8 8 8.788 8.864 8.783 8 6 8.728 8.828 8.781 8 6 8.728 8.828 8.484 8	202 040 406 1403 .444 6.242 6.022 6.026 6 .444 6.242 6.022 6.026 6 .404 6.232 6.206 6.210 6 .402 6.242 6.210 6.210 6 .204 6.446 6.214 6.414 6	1014 2220 1402 2010 1400 220 1.222 1.200 1.442 1.222 1.20 1.212 1.244 1.012 1.442 224 1.440 1.224 1.02 1.442 204 1.440 1.422 1.347 1.440 200 1.440 1.422 1.447 1.441	404 044 444 437 444 422 1.414 1.414 1.434 1.435 433 1.434 1.414 1.744 1.744 400 1.733 1.444 1.744 1.744 410 1.733 1.444 1.744 1.745 1.746 4.745 1.745 1.744 1.745 1.745	1 100 1200 1036 1036 1703 4 1.003 1.036 1.030 1.237 4 1.003 1.036 1.030 1.237 4 1.013 1.277 1.240 1.240 1 1.279 1.243 1.274 1.240 1 1.279 1.251 1.240 1.240	MA 241 441 641 1.422 1.441 1.441 1.442 1 1.422 1.441 1.441 1.442 1 1.422 1.441 1.441 1.442 1 1.421 1.441 1.441 1.442 1 1.421 1.441 1.441 1.441 1.442 1.442 1.444 1.441 1.441 1.441 1.442 1.441 1.441 1.441 1.441 1.444 1.441 1.441 1.441 1.441	144 121 144 214 14 1403 1423 1441 1424 144 1403 1424 1441 1424 1424 1424 1423 1424 1424 1424 1424 1424 1424 1424 1424	0 1024 401 10022 14400 62 0 1.001 1.024 1.406 1.461 1.3 4 1.222 1.442 1.342 1.402 1. 6 1.246 1.421 1.342 1.402 1. 6 1.246 1.401 1.342 1.402 1. 7 1.211 1.212 1.34 1.464 1.	146 (1483 6078 2068 (743) 2446 166 6.727 6.373 6.00 6.022 6.04 168 6.744 6.396 6.00 6.02 6.04 168 6.04 6.396 6.00 6.46 6.46 168 6.03 6.00 6.06 6.46 6.42 6.53 169 6.023 6.00 6.06 6.42 6.43 13 6.43 6.00 6.00 6.42 6.43	16430 1744 6471 6431 2413 2413 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1341 1343 1342 1342 1342 1342 1342 1342 1343 1343 1343 1342 1342 1342 1343 1343 1343 1343 1343 1343 1343 1343 1343 1343 1343 1343 1343 1343 1344 1343 <t< td=""><td>40 (144) 1231 (164) 4278 4283 2040 2014 201 201 1231 12 16 (201 2014) 1240 1238 1251 1213 121 1255 1240 120 40 (245 1252 1240 1244 1210 1210 1244 1242 1243 1244 14 (251 1252 1241 1214) 1254 1244 1257 1257 1257 1257 15 (251 124) 1251 1261 1255 1254 124 1257 1257 1257 1257 15 (251 124) 1251 1261 1257 1254 124 1257 1257 1257 15 (251 124) 1251 1261 1257 1254 124 1257 1257 1257 1257</td></t<>	40 (144) 1231 (164) 4278 4283 2040 2014 201 201 1231 12 16 (201 2014) 1240 1238 1251 1213 121 1255 1240 120 40 (245 1252 1240 1244 1210 1210 1244 1242 1243 1244 14 (251 1252 1241 1214) 1254 1244 1257 1257 1257 1257 15 (251 124) 1251 1261 1255 1254 124 1257 1257 1257 1257 15 (251 124) 1251 1261 1257 1254 124 1257 1257 1257 15 (251 124) 1251 1261 1257 1254 124 1257 1257 1257 1257
-	ALERS Fragm from Cr. Apolon (1.33) (1. ALERS Fragm from Cr. Abbott (1.13) (1.0) ALERS Fragm from Cr. Albott (1.13) (1.0) ALERS Fragm from Cr. Albott (1.33) (1.3) ALERS Fragm from S. Control (1.33) (1.3)	.4 68-86 8.862 88-84 8.84 81 8.138 8.824 8.876 8.83 81 34.88 8.823 14.31 4.73 24 87% 88% 88% 88% 732 21 84474 84478 84381 8438	4 1.710 1.400 1.343 1. 1 1.404 1.400 1.400 1. 4 1.443 1.440 2.433 1 5 74% 00% 74% 2 84370 84087 84383 84	.437 I.B.4 I.364 I.483 I 1484 I.434 I.414 I.488 I 1481 I.444 I.468 I.488 I 1492 I.444 I.468 I 1492 II.445 II.4936 II.494 1493 II.4413 II.4936 II.4956 II.	1-00 0.040 0.126 00-07 00-33 1 .106 0.006 0.031 0.123 0.43 1 1.60 1.04 3.444 33.31 003 26% 73% 73% 04% 66% 66%	144 2144 1.02 0.120 0.4 .02 0.00 0.02 0.020 0.04 4.04 0.20 0.274 2.440 00.0 02 248 0.8 248 00 044 22420 2048 2200 0240	4 1.344 1.046 1.447 00-00 6 0.01 0.022 30-00 0.007 1 1.062 6.140 0.000 14.47 6 01X 20X 00X 66X 4 03446 04414 04223 03740	10-02 20-14 00-10 0.007 1 1.220 0.224 0.420 0.021 1 220.7 62.26 74.00 0.721 1 172 912 002 022 022 12040 02206 01210 01440 1	1.440 1.114 1.427 20.04 1.11 10.00 1.02 1.004 1.070 1.04 1.004 2.74 1.014 14.02 1.00 1442 772 002 702 702 1430 14324 14000 12402 1250	4 8.444 8.482 88.84 8.482 8.3 6 68.84 8.888 8.488 8.488 8.4 2 8.66 8.498 8.489 8.49 8.4 2 888 8.48 848 848 848 848 8 888 848 848 848 848	20 0.706 0.006 0.403 0.064 3049 03 0.013 0.000 0040 0.034 0.403 00 0.06 00.4 0.040 0.010 00.03 08 042 332 002 002 603 04 040 0446 0446 0446 0446	20-04 0.00 40-04 0.424 0.002 40 0.200 0.000 0.004 0.004 0.002 0.0 24.7 2.242 0.00 0.440 0.0 14 442 172 222 202 012 172 2 7406 0406 0406 0406 0406 0406 04	10 10 10 10 10 10 10 10 10 10 10 10 10 1
	AICOS Compo Poets Holico 3000 3400 AICOS Compo Poets Adios 3043 344 AICOS Compo Poets Adios 1044 107 AICOS Compo Poets Adala 104 40 43 AICOS Compo Poets Monor 6,452 6,40 AICOS Compo Poets Monor 6,453 6,40	14 32733 30661 30306 3170 16 8438 4734 6378 438 13 1824 1784 3184 818 17 864 443 868 367 13 8.413 8.483 8.481 8.7 13 8.413 8.483 8.481 8.7	4 33868 31833 31488 3 6 7411 8248 7883 1 3 2448 1014 3868 3 7 3114 887 1338 4 8.843 8.413 8.831 8 6 8.287 8.884 8.244 8	1866 38668 33264 33249 32 8726 8883 8248 2883 9 8424 4683 3866 4343 3 864 4683 3866 4343 3 864 866 1426 3867 8 824 8.464 8.842 8.812 8	1788 31388 33363 1836 4143 8873 1418 8873 4883 1786 3388 3874 13143 3334 1348 6.778 6.888 1343 6.888 1348 6.778 6.888 6.444 6.888	1482 14827 11288 18773 3817 4718 6826 4371 8876 642 386 2777 888 1324 348 738 884 642 837 884 1324 738 884 644 643 434 884 743 8.848 6.444 8.848 8.848 8.844	33848 33333 33888 10483 4/38 3383 3384 3383 13848 4/38 3884 3383 13848 13848 5/38 3887 3887 3385 1884 6/38 3187 3187 3385 1884 6/38 3187 1271 3887 1894 1894 6/38 3187 1287 1297 3887 1294	31673 23348 31431 31837 3 4631 3014 016 2301 2346 (472 004 300 1030 074 641 1034 1.432 0.441 0.414 0.657	2322 31440 31631 33344 3004 2308 10004 6211 4440 844 1608 3102 (440 3006 340 844 1226 844 1068 164 1408 6.23 6.401 6.86 6.84 1.400 6.82 6.401 6.84 6.84	II 31873 38387 3 8344 8344 4 8444 3768 4 1888 1441 1883 3888 188 5 8.884 8.843 8.848 8.44 8.3	127 30300 10200 440 30400 32333 140 8,737 8,30 8,136 8,643 8,644 141 8,747 8,31 8,136 8,643 8,644	24422 8228 8828 8828 8827 6644 88 6.844 8.228 8.288 8.288 8.388 8. 6.442 8.666 8.222 8.2 8.2 8.2 8.2	
	2003 Tanja Parts Adam T.A.A. 1.0 2003 Tanja Parts Maan T.2018 1.25 2003 Tanja Parts Mala T.2018 1.25 2003 Tanja Parts Ajala 1.313 1.34 2003 Tanja Parts Ajala 1.313 1.34	14 1.030 1.00 1.023 1.40 A 1.734 1.76 1.733 1.00 A 40 40 40 40 40 M 40-06 1.044 1.00 1.40 A 1.007 1.027 1.007 1.00	4 6.231 6.00 6.20 6 4 6.003 6.20 6.421 6 5 40 40 40 4 6.403 6.200 6.430 1 5 60-04 6.001 6.002 6	.700 0.703 0.731 0.730 0 .743 0.404 0.400 0.404 1 40 40 40 40 40 0.711 0.43 0.376 0.371 0 .003 0.407 0.400 0.400 0	.201 0.422 0.404 0.204 0.442 1 0.42 0.420 0.02 0.024 0.421 1 40 40 40 40 40 40 0.16 0.022 0.004 0.40 20-44 1 0.16 0.006 0.442 0.001 20-44 1	.244 0.446 0.002 0.222 0.266 .203 0.406 0.223 0.402 0.40 40 40 40 40 40 40 0.04 20-04 0.042 0.023 0.040 .021 0.002 0.003 0.020 0.024	2 1.200 1.212 1.240 1.401 1 1.402 1.4 1.424 1.400 1 40 40 40 40 1 1.442 1.422 1.440 40 2 1.442 1.422 1.440 6.004	1.024 1.020 1.020 1.020 1.024 1.044 1.46 1.201 1.029 1 40 40 40 40 40 10.06 2042 2040 1.02 1.09 1.201 1.294 1.024	1.11 1.212 1.313 1.231 1.2 1.222 1.343 1.246 1.322 1.34 40 40 41 41 40 4 1.42 1.341 1.222 1.344 1.32 1.44 1.314 1.322 1.344 1.32	4 6.26 6.002 6.242 6.202 6. 2 6.222 6.223 6.236 6.264 6 8 40 40 40 40 40 4 6.064 6.412 6040 6.000 6. 2 6.000 6.012 6.064 6.040 6.	20 0.031 0.061 0.062 0.413 0.403 13 0.340 0.0.037 0.331 0.07 40 40 40 40 40 40 20 0.274 0.001 0.340 0.304 0.00 10 0.003 0.006 0.440 0.013 0.044	1.311 1.321 1.331 <th< td=""><td>23 1.37 1.021 1.021 1.012 1.014 1.02 1.014 1.01 1.014 1.01 (2) 1.344 1.014 1.013 1.014 1.015 1.016 1.014 1.014 1.014 1.211 1.211 1.014 1.014 1.014 1.01 1.015 1.021 1.014 1.014 1.211 1.211 1.014 1.014 1.014 1.01 1.014 1.014 1.014 1.014 1.317 1.011 1.012 1.012 1.014 1.014 1.014 1.014 101 1.014 1.015 1.013 1.015 1.014 1.014 1.014 1.014 1.014 101 1.014 1.015 1.014 1.015 1.014 1.014 1.014 1.014 1.014 101 1.014 1.015 1.014 1.015 1.015 1.014 1.014 1.014 1.014 101 1.014 1.015 1.014 1.015 1.015 1.014 1.014 1.014 1.014 1.014 101 1.014 1.014 1.014 1.015 1.015 1.014</td></th<>	23 1.37 1.021 1.021 1.012 1.014 1.02 1.014 1.01 1.014 1.01 (2) 1.344 1.014 1.013 1.014 1.015 1.016 1.014 1.014 1.014 1.211 1.211 1.014 1.014 1.014 1.01 1.015 1.021 1.014 1.014 1.211 1.211 1.014 1.014 1.014 1.01 1.014 1.014 1.014 1.014 1.317 1.011 1.012 1.012 1.014 1.014 1.014 1.014 101 1.014 1.015 1.013 1.015 1.014 1.014 1.014 1.014 1.014 101 1.014 1.015 1.014 1.015 1.014 1.014 1.014 1.014 1.014 101 1.014 1.015 1.014 1.015 1.015 1.014 1.014 1.014 1.014 101 1.014 1.015 1.014 1.015 1.015 1.014 1.014 1.014 1.014 1.014 101 1.014 1.014 1.014 1.015 1.015 1.014
	ALCUS - Except forbox - Moneto - 402 - 433 ALCUS - Except forbox - Const of - 14416 - 1440	X 40X 04X 04X 04X 703 13 14443 14420 14327 14343 13 4140 4066 6406 0103 16 7073 2000 0030 2007 10 1703 1207 3474 347	C 202 002 002 14326 14306 14364 14 6 6000 0402 6424 2 7 2404 2231 2460 2 6 3041 1400 3010 2	88% 28% 29% 29% 1423 14481 14341 14386 10 2886 4238 6231 6888 1 2483 4836 5281 2867 2 2669 28271 3184 3213 2	23% 64% 24% 86% 66% 1883 1488 1848 14448 14881 1 1484 4212 1846 18381 6481 4788 2124 2288	01% 21% 06% 22% 013 1003 13244 13443 13644 1424 1300 4400 2004 0447 044 1323 0140 4100 4000 003 1640 3234 1043 3133 344	C 882 282 282 842 1988 1942 1926 1917 2668 8268 6271 889 2967 8128 8276 887 2967 8128 8874 6446 2288 2968 2868 868	062 042 082 282 14140 14112 18223 18827 1 8187 8661 2883 6424 2486 8844 2238 2422 1223 1888 3883 3126	83X 73X 83X 76X 78 1938 19313 19346 19388 1913 2083 6343 7868 6084 64 2046 7331 7811 7634 733 3088 3869 3713 4018 331	X 74X 84X 38X 66X 3 8 19168 19443 19444 19444 19 8 7837 6876 13683 8881 18 6 7373 8987 7 3138 1988	8% 88% 31% 11% 87% 78% Na 1949a 1949a 1949a 1949a Al 1927 18997 18186 4681 6734	BIX 34X 38X 31X 13X 4 1949a 1949a 1949a 1949a 1949a 194 18383 18738 18838 18874 19493 18	42 012 132 12 112 112 132 132 132 132 135 145 145 132 147 148 145 145 145 145 145 145 145 145 145 145
-	AIGUS Comparisation Edular AIGUS Comparisation Monore C.A.W AIGUS Comparisation Tyrpon E.A.W AIGUS Comparisation Wolfer AIGUS Comparisation Molece C.A.W AIGUS Comparisation Molece C.A.W AIGUS Comparisation Molece C.A.W	4 237 278 383 113 1 6.466 6.46 6.448 6.84 14 6.488 6.86 6.448 6.84 14 6.488 6.88 6.84 15 6.88 6.88 6.84 15 6.81 6.82 14 6.817 6.82 14 6.84 1.48 6.82 1.48 6.84 1.4	4 070 204 404 0 0.064 0.402 0.422 0 0 0.244 0.004 0.004 0 0 0.216 0.064 0.20 0 0 0.420 0.044 0.206 1 1 0.400 0.222 0.042 0	103 624 626 664 674 200 6206 6266 6231 7 200 6246 6223 6231 7 206 6246 6223 6231 8 206 6203 6233 6231 8 620 620 620 621 621 8 204 620 621 621 621 6	041 1434 021 4052 1004 1.84 1.220 1.840 1.452 1.084 223 1.560 1.243 1.054 1.552 200 1.660 1.245 1.054 1.553 200 1.684 1.544 1.551 1.540 1.551 1.556 1.554 1.557 1.540	242 667 243 434 63 824 6.83 6.86 6.834 6.44 .844 6.244 6.86 6.252 6.86 244 6.247 6.86 6.231 6.23 248 6.652 6.84 6.247 6.23 248 6.65 6.231 6.44 6.245 6.2	4 838 964 708 884 9 8.966 8.949 8 8.839 9 8.948 8.949 8 8.839 9 8.948 8.925 8.939 8.639 8 8.948 8.638 8.638 8.649 8 8.644 8.648 8.648 8.648	420 204 240 440 0.431 0.444 0.440 0.024 0 0.46 0.443 0.012 0.291 0.46 0.022 0.463 0.293 1.740 0.023 0.231 0.400 0.240 0.023 0.231 0.400	347 051 314 440 54 1.844 0.814 0.818 0.834 0.85 1.831 0.734 0.837 0.753 0.75 1.748 0.846 0.741 0.713 0.73 1.741 0.842 0.731 0.840 0.84 1.741 0.842 0.731 0.840 0.84	2 612 894 684 814 814 4 4 6.842 6.466 6.804 6.842 6.3 8 6.244 6.443 6.229 6.666 6.3 8 6.245 6.816 6.223 6.662 6.3 9 6.824 6.234 6.35 6.866 6.2	N2 (N217 2447 1200 4042 1270) 40 0.641 0.24 0.042 0.666 0.664 12 0.647 0.218 0.07 0.622 0.7 10 0.441 0.446 0.66 0.60 0.622 10 0.44 0.621 0.642 0.442 0.200 10 0.244 0.621 0.642 0.247 0.50	4341 3274 3434 3431 1831 71 6.43 6.331 6.441 6.312 6.345	44 (12) 3447 (11) 2310 3471 (14) 1101 314 (12) 4(14) 1424 1544 (14) 1441 1541 1541 1531 (15) 145 145 27 1444 1571 1547 1541 155 1553 1553 1547 1547 154 16 1441 1577 1547 1553 1553 1544 1547 1548 1547 154 16 1557 1557 1557 1557 1557 1548 1547 1548 1547 15 1557 1557 1557 1558 1557 1548 1547 1548 1547 15 1557 1557 1557 1558 1557 1558 1557 1548 1548 1547 15 1557 1557 1558 1558 1557 1558 1557 1557
-	2003 - Corpo Indian Aprilan - B.400 - B.40 2003 - Corpo Indian - Dilati - B.40 - B.40 2003 - Corpo Indian - Chinai - B.442 - B.30 2003 - Corpo Indian - March - 402 - 432 2003 - Dorgon Indian - Canatal - 8330 - 834	17 20-12 0.01 0.02 0.01 12 0.416 0.066 0.004 0.021 14 00.42 4.200 0.02 0.25 14 00.42 4.200 0.02 0.24 2 00% 00% 00% 00% 742 12 02424 02204 02231 02131	0 0.003 0.303 0.464 0 0 0.034 0.010 0.006 3 6 3.409 3.030 0.010 0 6 00% 90% 01% 1 03271 01760 03036 0	.474 0.170 0.114 0.467 0 046 0.020 0.021 00-04 0 .042 0.400 4.24 0.066 1 042 702 748 022 1421 02217 02207 02167 0	.001 0.402 0.303 40-02 40-40 1 .044 30-04 0.033 0.044 1.62 3.71 0.036 3.000 37.40 303.0 70% 70% 00% 03% 63% 64%	1-04 61-06 8.821 8.822 20-0 .127 8.202 8.803 8.861 8.0 6.67 24.84 8.248 6.88 28.7 842 282 862 882 882 813 9448 84884 82226 44828 8167	6 1.762 1.740 1.000 6746 6 1.004 1.742 1.002 1.72 1 1.842 6.86 1.214 24.17 6 88% 28% 88% 67% 6 88% 28% 88% 67%	20-46 (0-00 20-00 0.70 1.002 0.204 0.442 0.004 204.6 26.04 67.27 0.447 102 402 072 022 0.22 1024 0006 20042 2014	0.43 0.026 0.320 20.04 0.02 0.001 0.002 0.021 0.00 0.04 0.140 2.324 3.040 14.60 6.70 04% 20% 00% 24% 00 12124 020% 00% 24% 000	4 0.222 0.400 0.004 0040 0.3 4 0.000 0.01 0.022 0.323 0.3 3 0.430 1.434 0.00 0.324 0.3 2 012 002 002 002 0.442 0 3 012 0245 03200 03200 032	06 0.00 0.003 0.440 30-04 00-03 22 0.036 0.006 30-00 0.027 0.300 0.3 0.004 03.00 0.000 06.40 36.30 42 022 302 0.300 00.80 06 10 10210 0217 02207 02207 02207	20-00 20-04 0.200 0.044 0.002 0.0 0.244 0.022 0.014 0.026 0.062 0.0 24.44 0.022 0.014 0.026 0.062 0.0 444 0.020 0.028 0.02 0.02 0.0 0200 0.0200 0.0200 0.0200 0.0200 0.02	23 23 23 23 23 23 23 23 23 23 23 23 23 2
-	ACCIS Compo Techen Medica (ASSA 3144 ACCIS Compo Techen Activa (ASSA 3144 ACCIS Compo Techen Medica (ASSA 114 ACCIS Compo Techen Activa (ASSA 214 ACCIS Compo Techen Activa (ASSA) ACCIS Compo Techen Activa (ASSA)	14 22024 20042 27044 2277 17 7420 7171 4704 1264 13 2004 2018 2204 702 10 1274 1404 1246 400 11 0.466 0.466 0.041 1 0.466 0.466 0.440 0.041	5 34004 30440 33134 33 3 11630 7010 11433 10 9 8347 3237 8343 3 9 3317 883 1434 1 1 8.874 8.483 8.433 8 1 8.747 8.848 8.414 8	2004 2077 22420 22460 22 0660 2220 2249 10840 22 0670 2202 4402 6073 4 1022 2470 2402 6073 4 1022 2470 2244 6402 6417 6412 6466 6417 1620 6570 6582 658	NAN 23031 23340 NAN 13431 13343 1827 2434 8004 1827 2434 8004 N33 4710 2412 2064 804 6.41 8.413 8.423 1 8.004 1 1.41 8.413 8.424 8.614	1294 31012 1609 10031 3243 682 0241 6301 0033 1002 3001 4006 1234 1046 129 1000 3221 421 1435 344 402 1.064 1.433 0.444 402 1.064 1.433 0.000 1.000 1.0	1 22604 24002 24400 20204 4 4440 0426 0062 0000 1 2200 0427 0021 2620 1 2400 0427 0021 2620 1 2400 0444 1 0.000 1 0.466 0.444 1 0.000	22414 22762 22267 21082 2 7001 6146 8284 1027 2020 2240 2243 2008 1616 407 100 1646 1.464 8.402 8.448 8.434 1 1.476 8.402 8.448 8.434 1	2327 21484 22146 22466 2126 18481 14747 18144 14871 1226 2887 4664 2884 2814 448 1228 2827 1284 1628 247 1428 8.428 8.487 8.41 8.42	0 21200 22204 4 12104 2214 5 2620 4210 2 2410 2202 20020 40040 16 2 0.400 0.466 0.000 0.047 0.3	231 44304 16414 2310 30340 43603 461 0.232 0.404 0.042 0.234 0.644 43 0.043 0.244 0.013 0.417 0.014	22167 12010 12200 12626 11204 24 642 6.0 6.60 6.607 6.207 6.4 6.426 6.00 6.407 6.207 6.40 6.	
-	ATERS Traja Talan Mahar) 1.024 1.0 ATERS Traja Talan Mahar) 1.025 1.0 ATERS Traja Talan Mahar 1.015 1.0 ATERS Traja Talan Mahar 1.016 10 ATERS Traja Talan Talan Talan 1.017 10 ATERS Traja Talan Talan 1.017 1.017	14 1.006 1.066 1.026 1.640 14 1.024 1.027 1.742 1.600 10 1.641 1.012 1.022 1.4 14 1.041 1.012 1.022 1.4 14 1.041 1.021 1.012 1.7 14 1.020 1.024 1.040 1.002	1 1.264 1.824 1.214 1 1 1.214 1.812 1.244 1 1 1.8 1.248 1.421 1 1 104 104 104 1 104 104 104 1 104 1.846 1.821 1 2 2040 1.822 4044 1	LUG E.200 E.255 E.214 E 2015 E.400 E.223 E.224 E 2023 E.101 E.112 E.154 E 104 E.101 E.112 E.154 E 204 E.240 E.246 E.246 E 246 E.240 E.246 E.246 E	246 1.440 1.26 1.246 1.400 234 1.44 1.221 1.401 1.402 429 1.2 1.421 1.401 1.402 109 109 109 109 109 100 1.400 1040 2040 2040 1.00 1.400 1.040 2040 2040	.00 0.20 0.004 0.220 0.20 240 0.400 0.002 0.240 0.20 442 0.402 0.440 0.4 0. 106 0.6 0.6 0.6 0. 0.6 0.60 0.640 0.200 4.0 444 0.600 0.600 0.200 0.00	2 1.224 1.246 1.231 1.631 1 1.246 1.668 1.233 1.683 1 1.204 1.688 1.233 1.683 1 1.204 1.649 1.646 1.239 4 1.04 1.04 1.04 1.04 1.233 1.633 1.643 4.644 1.613 1.633 1.644 1.600	1.002 0.000 0.040 0.240 0 0.020 0.002 0.026 0.204 0 0.0 0.46 0.402 0.024 0 00 00 00 00 00-00 00 00 00-00 0.026 0.025 0 0.000 0.026 0.02 0.025 0	LU22 1.746 1.82 1.717 1.77 1.747 1.718 1.747 1.724 1.73 1.667 1.718 1.747 1.724 1.73 1.667 1.723 1.143 1.143 1.144 104 104 104 104 10 1.438 1.701 1.713 1.104 1.33 1.438 1.701 1.132 1.132	4 6.264 6.020 6.220 6.240 2 6.26 6.240 6.24 6.242 6. 2 6.41 6.062 6.30 6.244 8 60 60 60 60 60 70 2 6.262 6.200 6.064 6.402 6. 6 6.00 6.01 6.20 6.01 6.0	1.3 1.031 1.746 1.740 1.441 1.310 170 1.431 1.716 1.721 1.411 1.312 18 1.431 1.716 1.721 1.411 1.313 18 1.741 1.613 1.746 1.314 1.613 18 1.741 1.614 1.746 1.316 1.533 24 1.717 1.717 1.4146 1.714 1.613	1.200 1.40 1.202 1.022 1.40 1.2 1.225 1.409 1.400 1.425 1.405 1. 1.422 1.422 1.400 1.423 1.406 1. 1.02 1.422 1.400 1.423 1.400 1. 1.03 1.400 1.200 1.400 1.401 1. 1.400 1.420 1.400 1.402 1.41 1.	14 1.244 1.247 1.248 1.249 1.249 1.249 1.249 1.248 1.247 1.248 1.2 1.244 1.248 1.244 1.249 1.252 1.244 1.246 1.247 1.246 1.27 1.2 1.244 1.247 1.246 1.247 1.246 1.247 1.248 1.248 1.247 1.4 1.44 1.244 1.248 1.249 1.249 1.248 1.248 1.248 1.248 1.25 1.251 1.248 1.248 1.248 1.248 1.248 1.248 1.248 1.25 1.251 1.248 1.248 1.248 1.248 1.248 1.248 1.248 1.25 1.251 1.248 1.248 1.248 1.248 1.248 1.248 1.248 1.248
	2003 Anno Ingilo Monte 172 40	× 78× 81× 73× 763	· 28X 81X 83X	NX 78X 79X 73X	172 112 212 412 112	11 X 11X 11X 11X 243	. 11 × 76× 11× 71×	BX 78X 64X 78X	122 112 122 142 14	X 77X 81X 34X 48X 1	3X MIX 13X 13X 43X MIX	38X 88X 38X 39X 8X 8	NX 21X 17X 6X 11X 4X 2X 8X 88X 84X 73;

My own development of an automated reporting engine began with the need to handle substantial amounts of information for a single large institution. There was a lot of data to get through!

- .
- 7

	Overall				
Innovative or creative nas developed during my course	Q16_3_a My research skills have developed during my course	Q16_4_a My ability to communicate information effectively to diverse	Q16_5_a1 have been encouraged to think about what skills1 need to develop	Q16_6_a As a result of the course I feel better prepared for my future career	Q18_1_a Overall, I am satisfied with the quality of the course
74%	82%	73% (n=	=5561) is	78%	83%
77%	77% 73%		5501715		
		signincar	tly lower th	an 86%	81%
74%	73% 84%	82% (n=	=81742). Ch	i- 77%	86%
73%	79%	Square (=81742).Ch 2,87303) =	i⊦ 77% 84%	86% 85%
73%	79% 81%	Square (=81742). Ch	i- 77% phi 84% 81%	86% 85% 85%
73% 70% 69%	79% 81% 81%	Square (=81742).Ch 2,87303) =	i⊦ 77% phi 84% 81% 80%	86% 85% 85% 80%
73% 70% 69% 73%	79% 81% 81% 82%	Square (343.52, = 1.16	=81742). Ch 2,87303) = p = 0.000,	ii 77% 84% 81% 80% 85%	86% 85% 85% 80% 85%
73% 70% 69% 73% 73%	79% 81% 81% 82% 83%	Square (343.52, = 1.16	=81742). Ch 2,87303) = p = 0.000, p = 80%	i⊢ 77% 84% 81% 80% 85% 82%	86% 85% 80% 85% 86%
73% 70% 69% 73% 73% 73%	79% 81% 81% 82% 83% 86%	Square (343.52, = 1.16	=81742). Ch 2,87303) = p = 0.000, p = 80%	i⊢ 77% 84% 81% 80% 85% 82% 79%	86% 85% 85% 80% 85% 86% 83%
73% 70% 69% 73% 73%	79% 81% 81% 82% 83%	Square (343.52, = 1.16	=81742). Ch 2,87303) = p = 0.000, p = 80%	i⊢ 77% 84% 81% 80% 85% 82%	86% 85% 80% 85% 86%

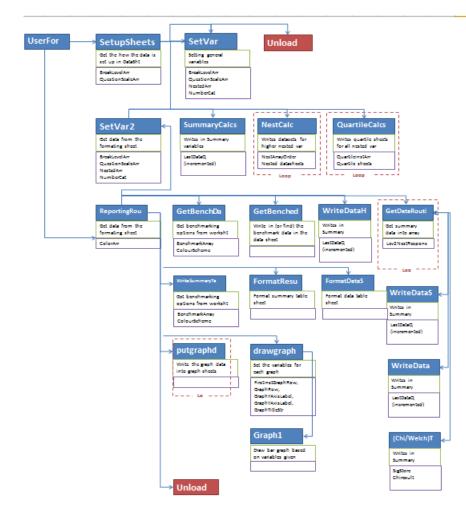
Initial reporting needs were driven by the need for multiple statistical tests across numerous breakdowns to identify where important differences exist.

Chi-square was used as a relatively robust test to identify important differences where response numbers varied widely.

However, significance levels of p<0.05 often produced too many flags, whilst p<0.01 too few. Development needed to incorporate more sophistication.



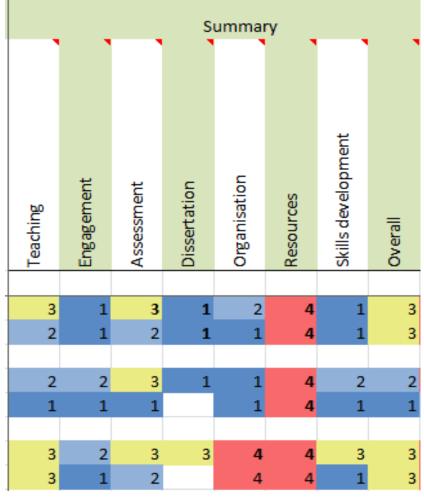
I then moved from a single institution to the Higher Education Academy, where reports were needed for over 150 institutions. This created an additional challenge.



The engine has been rewritten twice, with increasing flexibility, capability, and integration.

The engine is written in VBA, separate stages of reporting processes being handled by largely separate routines.

Currently the engine has 3747 lines of code, with six essentially separate modules covering the process from raw data to report production.



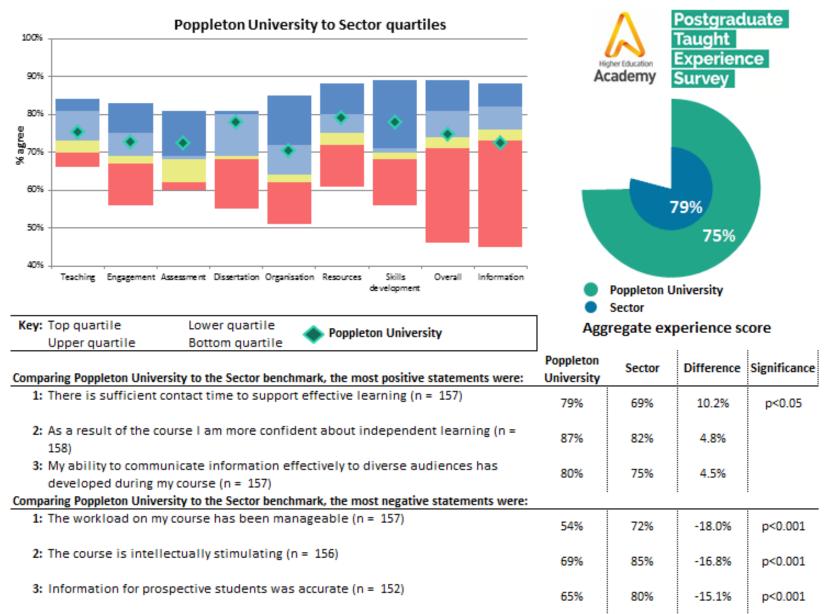
Quartile and weighted calculations are precalculated prior to reporting. Sheets can be filtered to produce quartile or effect size heat maps.

Effect sizes were used to define the colouring of significant cells. This meant that many significance flags were created, but prioritised through colour coding of cells according to effect size. The Holm– Bonferroni adjustment for multiple chi-squared tests was added to reduce false positives.

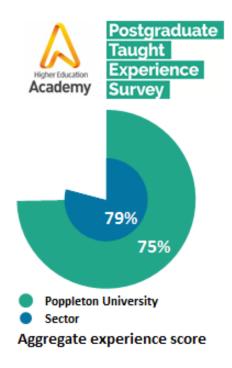
The expansion of statistical calculations
involved in the creation of each report,
along with the need for overviews of
statistical calculations, prompted a move
away from large arrays. Multiple temporary
sheets were used to store interim
calculations instead.

	А	С	D	E	F	G	Н	1	J	K	L	М	N
1			Summary										
2	● PTES 2016 results	Average Response	Teaching	Engagement	Assessment	Dissertation	Organisation	Resources	Skills development	Overall	Information		Q2_1_a Staff aregood at explaining things
3	All responses												
14	Gender												
15	Male	31684	82%	79%	72%	80%	75%	84%	79%	83%	84%		88%
16	Female	47384	83%	79%	74%	78%	74%	84%	77%	82%	85%		88%
17	Disability										78%	6 (n=63	363) is
18	Disabled	6149	79%	75%	71%	75%	69%	82%	73%	78%			lower than
19	No known disability	71140	83%	79%	74%	79%	74%	85%	78%	83%			2476). Chi-
20	Fee status												38839) = 0.000, phi
21	UK	38080	83%	79%	75%	79%	72%	84%	77%	83%	= 0		0.000, pill
22	Other European Union (EU)	10072	78%	75%	66%	75%	70%	83%	73%	78%	1		
23	Non-EU	31321	82%	79%	74%	80%	77%	85%	80%	83%	1		
24	Nationality										-		
25	UK incl. Channel Islands	41588	83%	79%	74%	79%	73%	84%	77%	83%	85%		88%
26	Asia	15410	84%	80%	77%	81%	79%	86%	82%	85%	86%		90%
27	Other Europe	1125	79%	75%	67%	73%	73%	84%	74%	79%	84%		86%
28	Africa	3568	87%	82%	77%	82%	81%	86%	87%	89%	90%		91%

Summary for Poppleton University compared to the Sector benchmark.



The quartile graph shows the range of institutional scores in PTES for each measure, divided into quartiles. For example, the lower quartile covers the range from the lowest scoring institution across a quarter of the lower scoring institutions. The diamond indicates where this institution is on the scale. The overall satisfaction graph shows the institution and benchmark scores for the overall question. The positive and negative statements are the three statements (where an agree/diagree scale is used) with the largest positive and negative differences.



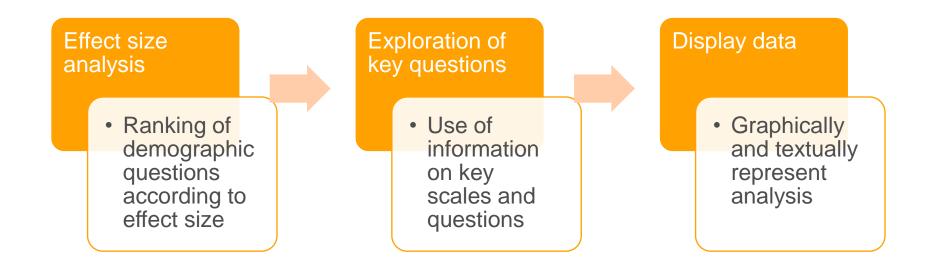
From summary tables the next step was a dashboard that sought to enable clear prioritisation of action.

This process begun with graphical representations of relative position.

It was then developed through textual descriptions of the largest differences and the listing of priority questions related to those differences.

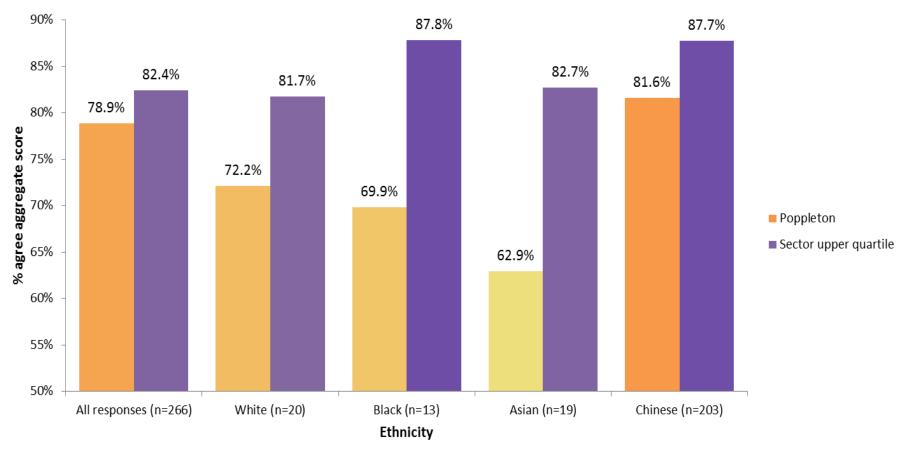
However, when it came to displaying demographic analyses, a method was needed to narrow down the many possible questions.

> Demographic analysis...



One challenge was to select particular demographic questions to display. This required a prioritisation. Using a research strategy of selecting questions of interest by effect size, the engine progresses through logical steps to present key analyses.

Difference between Poppleton and Sector upper quartile, by Ethnicity



There were relatively large differences within Ethnicity for poppleton compared to the Sector upper quartile benchmarking group. 'Chinese' had the least negative results at poppleton relative to the benchmark with an aggregate score -6.1% lower than the Sector upper quartile benchmark. This was a significant (p<0.05) difference. The category most negative relative to the Sector upper quartile benchmarking group was 'Asian', with a score 19.8% lower than the Sector upper quartile benchmark.

Within Ethnicity - Chinese, the most positive area of feedback was Dissertation. Here, the most positive statements were:	poppleton	Sector upper quartile	Difference	Significance					
1: I understand the required standards for the dissertation / major project	93%	89%	3.5%						
2: I am happy with the support I received for planning my dissertation / major project	86%	86%	0.1%						
Within Ethnicity - Asian, the most negative area of feedback was Resources. Here, the most negative statements were:									
1: The library resources and services are good enough for my needs	37%	87%	-49.7%	p<0.001					
2: I have been able to access subject specific resources necessary for my studies	53%	86%	-33.9%	p<0.001					
Graphical, numerical and textual presentation of data increases the potential that the end user will understand and disseminate findings, regardless of their understanding of statistical methods.									

Automated processes:

Customisation of reports around benchmarking groups, colour scheme, etc

Tabular statistically tested output

Analysis based upon difference, significance and effect size

Graphical and textual representation

Objective and comprehensive use of statistics to define priorities according to data

Non-automated process:

Aggregation of variables into grouped variables and scales

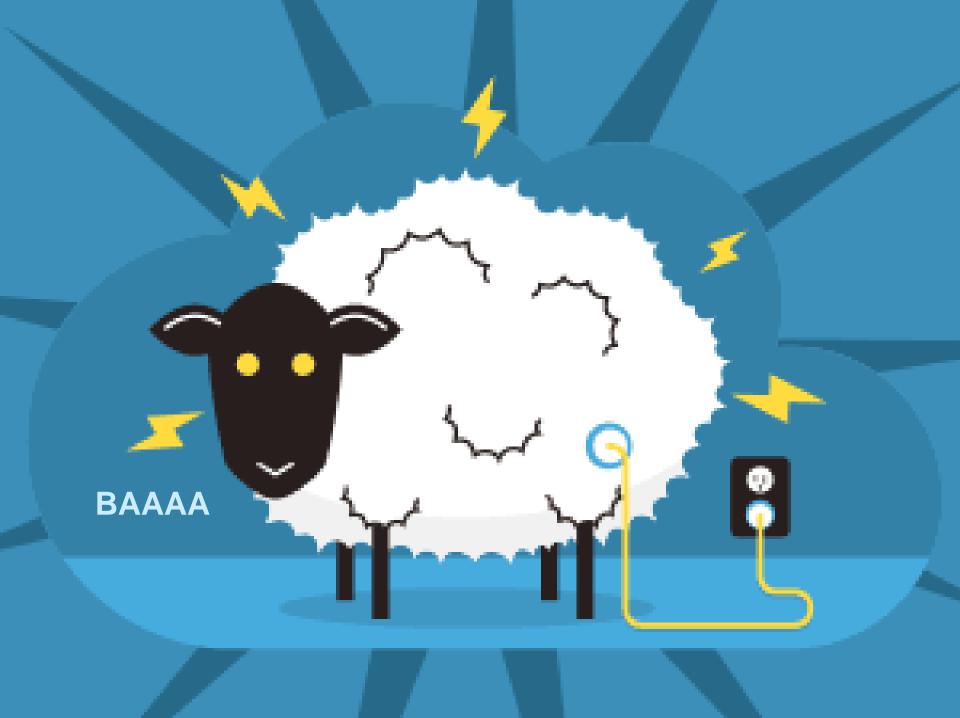
Design and format of analysis

Further alteration and adaptation

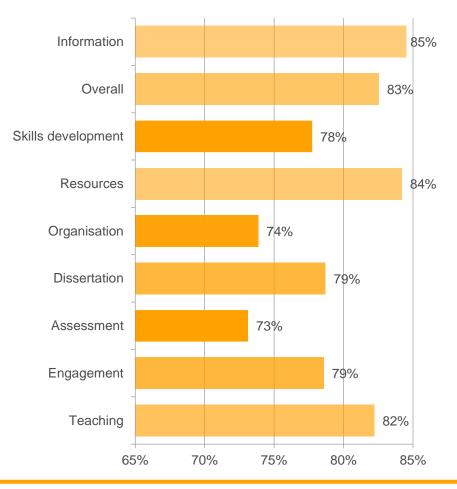
Links to enhancement and strategy – the issues and context of the institution

Subjective and partial use of statistics to define priorities according to strategy

The engine does not automate many functions, such as factor analysis. End users often wish to select particular parts of analyses or otherwise adapt to fit into institutional context and enable enhancement. The engine is being developed to improve how it informs users of research choices.



What analysis is reported is informed by issues and context.

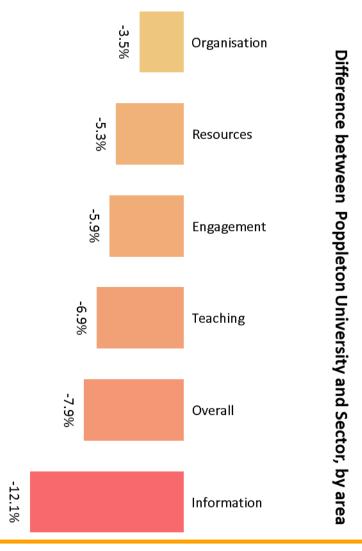


What is the priority – the lowest scoring scales such as Assessment and Organisation?...

Area

What analysis is reported is informed by issues and context.

How does this decision impact upon dissemination, enhancement, politics?

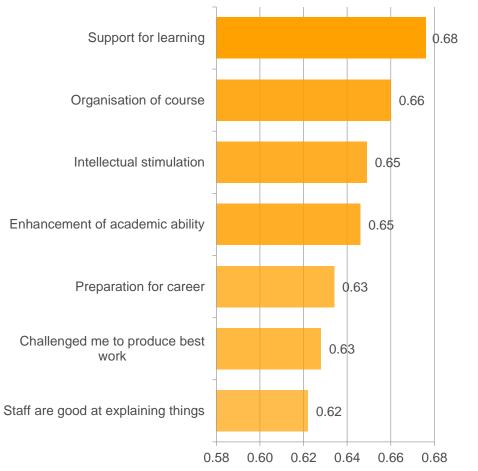


...Or is the priority the scale that scores lowest relative to other institutions?...

What analysis is reported is informed by issues and context.

How does this decision impact upon dissemination, enhancement, politics?

IR is moving up Terenzini's hierarchy, from technical analysis to issues and context.



...Or is the priority the area that is most correlated with the overall experience? Whilst automation can make this analysis easier, the choice cannot be automated.

What analysis is reported is informed by issues and context.

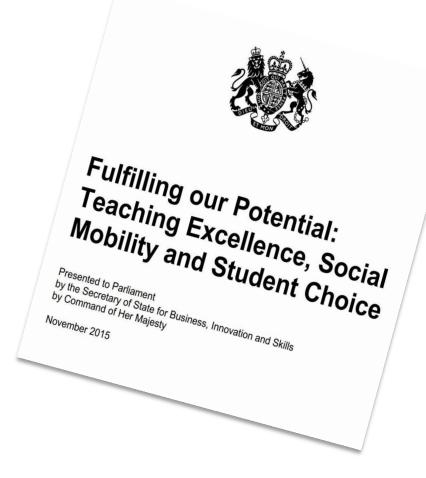
How does this decision impact upon dissemination, enhancement, politics?

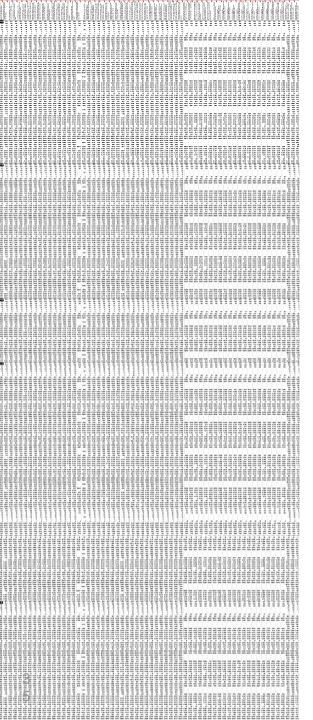
IR is moving up Terenzini's hierarchy, from technical analysis to issues and context.

IR connects to theory, strategy, and teaching & learning practice – across departments, institutions, and the sector.

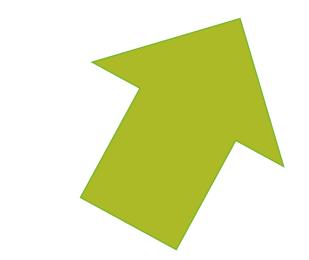
3

4





The value of automated analysis is in giving people more sophisticated tools to find differences that matter...



...the value of the institutional researcher is increasingly about bridging this analysis to the context and issues institutions face.

References

The Automatic Statistician http://www.automaticstatistician.com/

Frey, C. B., & Osborne, M. A. (2013). The future of employment: how susceptible are jobs to computerisation. <u>http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf</u>

Hammond, N. (2009) Institutional research and the student learning experience, Perspectives: Policy and Practice in Higher Education, 13:3, 80-83

King, B. R. (2016). Making the Connections Across Institutional Types and Academic Programs: Recommendations for Institutional Research Practice and Future Research. *New Directions for Institutional Research*, 2015(168), 101-105.

Volkwein, J. F. (1999). The four faces of institutional research. *New directions for institutional research*, *1999*(104), 9-19.

Wells, R. S., & Stage, F. K. (2015). Past, Present, and Future of Critical Quantitative Research in Higher Education. *New Directions for Institutional Research*, *2014*(163), 103-112.

Questions?

jason.leman@heacademy.ac.uk

@HEA_surveys